DOCUMENT RESUME

ED 467 370 TM 034 296

AUTHOR	Ben-Yashar, Ruth; Nitzan, Shmuel; Vos, Hans J.
TITLE	Optimal Cutoff Points in Single and Multiple Tests for Psychological and Educational Decision Making. Research Report.
INSTITUTION	Twente Univ., Enschede (Netherlands). Faculty of Educational Science and Technology.
REPORT NO	RR-01-04
PUB DATE	2001-00-00
NOTE	38p.
AVAILABLE FROM	Faculty of Educational Science and Technology, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands.
PUB TYPE	Reports - Research (143)
EDRS PRICE	EDRS Price MF01/PC02 Plus Postage.
DESCRIPTORS	*Cutting Scores; *Decision Making; *Educational Testing; *Personnel Selection; Prediction; Psychological Testing; *Test Use

ABSTRACT

This paper compares the determination of optimal cutoff points for single and multiple tests in the field of personnel selection. Decisional skills of predictor tests composing the multiple test are assumed to be endogenous variables that depend on the cutting points to be set. The main result specifies the condition that determines the relationship between the optimal cutoff points for single and multiple tests, given the number of predictor tests, the collective decision rule (aggregation procedure of predictor tests' recommendations) applied, and the function relating the tests' decisional skills to the cutoff point. The implications of the main result are developed for special types of collective decision rules such as disjunctive and conjunctive rules. The proposed group decision-making method is illustrated by an empirical example of selecting trainees for a large company by means of the assessment center method. How the predictor cutoffs and the collective decision rule are determined dependently by maximizing the multiple test's common expected utility is shown. The analysis suggests that centralized decision-making systems can improve their performance by changing the cutoff points determined separately for each screening predictor. Two appendixes present two theorems for decision making. (Contains 1 figure and 29 references.) (Author/SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 467 370

TM

Optimal Cutoff Points in Single and Multiple Tests for Psychological and Educational Decision Making

Research Report 01-04

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

J. Nelissen

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Ruth Ben-Yashar
Shuel Nitzan
Bar-Ilan University

Hans J. Vos

TM034296

faculty of
EDUCATIONAL SCIENCE AND TECHNOLOGY

University of Twent

Department of
Educational Measurement and Data Analysis

BEST COPY AVAILABLE

Optimal Cutoff Points in Single and Multiple Tests for Psychological and Educational Decision Making

Ruth Ben-Yashar

Shmuel Nitzan

Bar-Illan University
Department of Economics
52900 Ramat-Gan
Israel

Hans J. Vos

Abstract

This paper compares the determination of optimal cutoff points for single and multiple tests in the field of personnel selection. Decisional skills of predictor tests composing the multiple test are assumed to be endogenous variables that depend on the cutting points to be set. Our main result specifies the condition that determines the relationship between the optimal cutoff points for single and multiple tests, given the number of predictor tests, the collective decision rule (aggregation procedure of predictor tests' recommendations) applied and the function relating the tests' decisional skills to the cutoff point. The implications of our main result are developed for special types of collective decision rules such as disjunctive and conjunctive ones. The proposed group decision-making method is illustrated by an empirical example of selecting trainees for a large company by means of the assessment center method. It is shown how the predictor cutoffs and the collective decision rule are determined dependently by maximizing the multiple test's common expected utility. Our analysis implies that centralized decision-making systems can improve their performance by changing the cutoff points determined separately for each screening predictor.

Introduction

Over the past few decades, much psychometric research has been aimed at improving the use of educational and psychological tests as means for decision making rather than for estimating ability scores from test performances. Examples of such decisions are admittance of students to a university and personnel selection in industry (e.g., Chuang et al., 1981; Cronbach & Gleser, 1965; De Corte, 1998; Petersen, 1976; Raju et al., 1991; van der Linden & Vos, 1996; Vos, 1997a), pass-fail decisions in education and successfulness of therapies in psychodiagnostics (e.g., Huynh, 1977; Lewis & Sheehan, 1990; Vos, 2001), optimal assignment of students to different instructional treatments in Aptitude Treatment Interaction (ATI) research (e.g., Cronbach & Snow, 1977; van der Linden, 1981; Vos, 1997b), and vocational guidance decisions in which most promising schools or careers must be identified (e.g., van der Linden, 1987). Optimal cutoff points can be found by formalizing each of the above types of elementary test-based decisions as a problem of Bayesian decision making by maximizing its expected utility (e.g., DeGroot, 1970; Lehmann, 1959).

The existing psychological and educational literature discusses how cutoff points can be determined, while there is only one test or one measure which weighs the scores on a number of tests as a composite score, or, for many tests, how the cutoff point on each test can be determined separately. However, no results are reported how in case of a multiple test composed of several tests the cutoff points on each separate test and the collective rule (i.e., aggregation procedure) can be determined dependently. For example, take a predictor-based selection system in which the collective decision rule is that an applicant must pass $(n+1)/2$ out of n predictor tests, then one must decide on a cutoff point for each separate predictor test.

Therefore, the goal of this paper is to present a model that takes into account the dependence between the cutoff points on a number of predictor tests composing a multiple test and their aggregation process to come to a collective decision in terms of rejecting or admitting an applicant for a job in industrial/organizational (I/O) psychology. In other words, the cutoff points and the aggregation rule will be optimized simultaneously by maximizing the multiple test's common expected utility.

In essence, personnel selection involves the screening and comparison of applicants by means of valid procedures with the purpose of obtaining the intended quota of supposedly most successful employees (e.g., Cronbach & Gleser, 1965). In selecting the required quota or selection ratio (i.e., fixed proportion of all applicants that can be accepted due to shortage of resources), a mechanism of top-down selection is usually followed by selecting applicants with the highest scores

on the (composite) predictor test until the quota is filled. For instance, the well-known Taylor-Russell (1939) tables are based on this formalism.

In addition, top-down selection formalisms have been proposed in personnel selection that are based on the expected average criterion score of the selected applicants (e.g., Boudreau, 1991; Brogden, 1949; Cronbach & Gleser, 1965; De Corte, 1994). In fact, these methods are prevailing in the literature but cannot be employed if future criterion behavior (i.e., job performance) is assumed to be a dichotomous variable, that is, either successful or not (e.g., Raju et al., 1991). In these situations we have to resort to the expected success ratio, that is, the proportion of applicants accepted that will be successful in their future job performance.

Although the existing psychology literature dealing with classification procedures in terms of acceptance/rejection usually takes the given quota into account, this approach is not followed in the present paper. The main reason is that there are many problems in which it is not feasible to assume quota restrictions, for example, deciding on whether or not to hospitalize a patient. In the field of personnel selection, also situations may exist in which it is not feasible to impose quota restrictions. For instance, if we want to select all applicants who showed satisfactory performance on a fixed number of predictor tests. This is exactly the situation where the present paper is aiming at by selecting all applicants with scores on each predictor test higher than a cutoff point that clearly depends on the number of predictor tests composing the multiple test, the aggregation process of predictor tests' recommendations, and the function relating the tests' decisional skills to the cutoff point. In principle, unlike the fixed quota methods, none or all of the applicants might be selected with the unconstrained dichotomous choice model proposed in the present paper.

The problem addressed in the present paper shows some correspondence to the case of multiple hurdles or multi-stage selection in I/O psychology (e.g., De Corte, 1998, Milkovich & Boudreau, 1997; Sackett & Roth, 1996). The case of multiple hurdles deals with a situation in which an applicant is expected to show minimum proficiency in several skilll areas. In this scenario, as opposed to, for instance, multiple regression analysis, a high proficiency in one skill area will not typically compensate for a low proficiency in another skill area.

The model advocated here has been applied earlier successfully by Ben-Yashar and Nitzan (1997, 1998, 2001) to the field of economics where organizations face the comparable problem of deciding on approval or rejection of investment projects. A team of n decision makers has to decide which ones of a set of projects are to be accepted so as to maximize the team's common expected utility. The proposed group decision-making method can be applied to many binary decisions determined by teams of decision makers or test systems.

The paper is organized as follows. Section 2 applies the Ben-Yashar & Nitzan economic model to the field of personnel selection. In Section 3 we derive necessary and sufficient conditions for optimal cutoff points of single and multiple tests. In Section 4 the optimal cutoff points set on single and multiple tests are compared by deriving an inequality that specifies the relationship between these two types of cutoff points. Section 5 focuses on the comparison of the two types of cutoff points for special types of collective rules, namely, disjunctive and conjunctive ones. The comparison between optimal cutoff points for single and multiple tests in predictor-based selection is empirically illustrated in Section 6 where applicants are either accepted or rejected as trainees for a large company by means of the assessment center method. The concluding section contains a brief summary of the main result and discusses a few possible lines of future research by loosening some of the assumptions in the present framework.

The Model

In the field of personnel selection, it often occurs that an applicant is either accepted or rejected for a job based on a multiple test composed of several predictor tests, i.e., a battery of n ($n \geq 1$) performance measures such as psychological tests, role-plays, and work sample tasks. It is assumed that the true state of an applicant regarding the future job performance (usually a supervisory performance rating) is unknown and can be qualified as either suitable ($s = 1$) or unsuitable ($s = -1$). An applicant is qualified as suitable if his or her performance is at least equal to a pre-established cutoff point (performance level) on the criterion variable(s) represented by the future job performance. Furthermore, based on applicant's performance on predictor test i ($1 \leq i \leq n$), it is decided if an applicant is passed ($a_i = 1$) or failed ($a_i = -1$) on predictor test i. The decision table for each predictor test i is therefore:

Decision	**State Applicant**	
	Suitable	Unsuitable
Pass	(1,1)	(1,-1)
Fail	(-1,1)	(-1,-1)

The true state of an applicant, however, is unknown on each of the n predictor tests. Instead, an applicant receives a test score x_i (i.e., a performance rating) on each predictor test i which

depends on applicant's performance in a certain skill area. It is assumed that the scales of the predictor tests have been transformed such that they cover all the same range of test scores. The pass-fail decision a_i is now made by setting a cutoff point on each test score x_i in the form of a threshold R (i.e., predictor cutoff) such that

$$x_i \geq R \Rightarrow a_i = 1$$
$$x_i < R \Rightarrow a_i = -1.$$

The test score x_i is drawn from a known distribution function represented by the density $f_1(.)$ for suitable and $f_2(.)$ for unsuitable applicants. Therefore, the conditional probabilities p_i^1 and p_i^2 that a predictor test i makes a correct pass-fail decision under the two possible states of nature (the *decisional skills* of each predictor test) are:

$$p_i^1 = \Pr\{a_i = 1 \mid s = 1\} = \int_R^{\infty} f_1(x_i)\mathrm{d}x$$

and

$$p_i^2 = \Pr\{a_i = -1 \mid s = -1\} = \int_{-\infty}^{R} f_2(x_i)\mathrm{d}x,$$

where $(1- p_i^1)$ and $(1- p_i^2)$ can be interpreted as Type I and Type II errors (i.e., conditional probabilities of making incorrect pass and fail decisions) of each predictor test i. Decisional skills of predictor tests are assumed to be endogenous variables that depend on the cutoff points to be set. Our assumptions imply both decentralized information processing and limited communication; the information available of a predictor test i is transformed into a binary signal (pass/fail) and it is only these signals that are communicated.

In the sequel, although the predictor tests i will usually differ in their outcomes regarding passing or failing of applicants, it will be assumed that the decisional skills of each predictor test i are equal. As a consequence, the index i in p_i^1 and p_i^2 will be dropped from now on. In other words, $\forall i,\ p_i^1 = p_1,\ p_i^2 = p_2$. This assumption of so-called *homogeneous* decisional skills is realistic if, based on an applicant's proficiency in a certain area, the assessors (or teams of assessors) who must decide if an applicant is either passed or failed on a predictor test i are equally skilled.

It is clear that by increasing R, Type I error $(1\text{-}p_1)$ also increases, and Type II error $(1\text{-}p_2)$ decreases. Similarly, decreasing R reduces Type I error and increases Type II error. Since $\frac{dp_1}{dR} = -f_1(R) < 0$ and $\frac{dp_2}{dR} = f_2(R) > 0$, these requirements obviously hold.

Note that we make the following assumptions: (i) the predictor tests that usually differ in their outcomes are homogeneous in their decisional skills, (ii) $\frac{p_1 + p_2}{2} > 0.5$. This assumption implies that $p_1 > (1\text{-}p_2)$, that is, a suitable applicant is more likely to be passed on a predictor test than an unsuitable applicant. Thus, since the simple average of the test's decisional skills in the two states of nature exceeds 0.5, each predictor test participating in the collective decision-making process is valuable because this test's decisional skill is superior to that of a random decision process. (iii) decisional skills of predictor tests are statistically independent. It should be noted that the assumption of statistical or local independence is also frequently made in other applications of psychological and educational tests. Local independence then means that when the abilities influencing test performance are held constant (i.e., conditioning on ability), examinees' responses to any pair of items are statistically independent. In fact, local independence is one of the basic assumptions made in Item Response Theory (IRT) models (e.g., Hambleton et al., 1991).

The vector $a = (a_1, \ldots, a_n)$ is referred to as the *decision profile* of a set of n predictor tests for an individual applicant, where $a_i = 1$ or $a_i = -1$ denotes if the applicant is either passed or failed on predictor test i $(1 \le i \le n)$. The collective decision, acceptance (1) or rejection (-1) of an applicant, is then determined by means of a decisive aggregation rule g that transforms the profile of decisions on n predictor tests into a collective decision. g is referred to as the *structure* of the collective decision-making process and assigns 1 or −1 (acceptance or rejection of an applicant) to any decision profile a in $\Omega = \{1,\text{-}1\}^n$. That is, g: $\Omega \rightarrow \{1,\text{-}1\}$. The same problem is faced in the multiple hurdles scenario where, based on applicant's performance on several tests, a collective decision 1 or −1 (acceptance or rejection of an applicant) must be made.

To formally define the objective function (i.e., the multiple test's expected common utility), we need to present the conditional probabilities of reaching a correct collective decision, given the structure g. Let us therefore partition the set Ω of all decision profiles into $A(g/1)$ and $A(g/\text{-}1)$, where $A(g/1) = \{a \in \Omega \mid g(a) = 1\}$ and $A(g/\text{-}1) = \{a \in \Omega \mid g(a) = \text{-}1\}$,where $g(a)$ is the collective decision for a decision profile a. For a given structure g, the collective decision-making process accepts a suitable applicant and rejects an unsuitable applicant with probability $\varphi(g/1)$ and $\varphi(g/\text{-}1)$,

respectively, where $\varphi(g/1) = \Pr\{a \in A(g/1) \mid s = 1\}$ and $\varphi(g/-1) = \Pr\{a \in A(g/-1) \mid s = -1\}$. Note that for a single test, $\varphi(g/1)$ and $\varphi(g/-1)$ are equal to respectively p_1 and p_2.

Necessary and Sufficient Conditions for Optimal Cutoff Points

For a multiple test, our goal is to derive the collective decision rule g and cutoff point R dependently that maximize the multiple test's common expected utility. Therefore, the following problem is faced:

$$\begin{aligned}\text{Max}_{R,g}\ & \alpha U(1/1)\varphi(g/1) + \alpha U(-1/1)(1-\varphi(g/1)) + \\ & (1-\alpha)U(-1/-1)\varphi(g/-1) + (1-\alpha)U(1/-1)(1-\varphi(g/-1)),\end{aligned} \tag{1}$$

where $U(1/1)$, $U(1/-1)$, $U(-1/-1)$ and $U(-1/1)$ are the (economic) utilities corresponding to the four possible decision outcomes on each predictor test, that is, correct passing (true positive), incorrect passing (false positive), correct failing (true negative), and incorrect failing (false negative). Furthermore, $\alpha\ (\alpha \neq 0,1)$ and $(1-\alpha)$ denote the a priori probabilities that an applicant is qualified as either suitable (1) or unsuitable (-1). Since $[\alpha U(-1/1) + (1-\alpha)U(1/-1)]$ does not depend on R, the above maximization problem can be reduced to the following form:

$$\begin{aligned}\text{Max}_{R,g}\ & \alpha U(1/1)\varphi(g/1) - \alpha U(-1/1)\varphi(g/1) + \\ & (1-\alpha)U(-1/-1)\varphi(g/-1) - (1-\alpha)U(1/-1)\varphi(g/-1).\end{aligned}$$

Let $U(1) = [U(1/1) - U(-1/1)]$ denote the positive net utility corresponding to the correct pass decision, and let $U(-1) = [U(-1/-1) - U(1/-1)]$ denote the positive net utility corresponding to the correct fail decision, it then follows that the maximization problem in (1) can be formulated as:

$$\text{Max}_{R,g}\ \alpha U(1)\varphi(g/1) + (1-\alpha)U(-1)\varphi(g/-1). \tag{2}$$

Threshold utility

It should be mentioned that in fact a so-called threshold utility function is assumed in the present paper. That is, the utilities involved can be summarized by possibly different constants for each of

the four possible decision outcomes (i.e., fixed utilities). In other words, although the utilities depend indirectly on the value of the predictor cutoff R via the pass-fail decision, they do not explicitly depend on R. For instance, the utility corresponding to an incorrect pass decision on predictor test i (i.e., $U(1/\text{-}1)$) for an unsuitable applicant who is far above R is the same as for an incorrect pass decision for an unsuitable applicant who is performing just above R. This will be true for both suitable applicants and unsuitable ones. Considering the joint distribution of the applicant predictor and criterion scores, it is also obvious that the expected criterion score (and hence, the utility) of an applicant who passed predictor test i will vary for different cutoff values of R. Most current models of personnel selection utility, therefore, follow the classical Brogden-Cronbach-Gleser suggestion (Brogden, 1949; Cronbach & Gleser, 1965) to express the utility explicitly as a function of the predictor cutoff R.

However, like the model proposed in this paper, some models of personnel selection utility assume that utility does not explicitly depend on the value of R by adopting a threshold utility function (e.g., Chuang et al., 1981; Petersen, 1976; Raju et al. 1991; Vos, 2001). Threshold utilities are also frequently assumed as being appropriate in the context of educational decision making (e.g., Huynh, 1977; Lewis & Sheehan, 1990; van der Linden, 1987). The main reason for defending threshold utility by all these authors is that, referring to the previous given example, applicants with scores on predictor test i far above R will hardly never be qualified as unsuitable. Moreover, these authors assume that utilities corresponding to the correct pass and fail decisions (i.e., $U(1/1)$ and $U(\text{-}1/\text{-}1)$) remain relatively stable for applicants with predictor scores respectively far above and far below R. So, if it is assumed that utilities are in fact only sensitive to changes in predictor scores around the cutoff point R, the discontinuous threshold function as a "jump" from one constant value to another can be defended as a realistic model for personnel selection utility.

Finally, it can still be remarked that threshold utilities are quite convenient from a mathematical point of view. As will become clear below, not the absolute utilities $U(1/1)$, $U(\text{-}1/1)$, $U(\text{-}1/\text{-}1)$ and $U(1/\text{-}1)$ have to be specified for computing the optimal cutoff points but only the so-called utility ratio $U(1)/U(\text{-}1)$, that is, $U(1)$ relative to $U(\text{-}1)$, has to be specified.

Most texts on decision theory propose lottery methods for empirically assessing the fixed values of the threshold utility function (and hence, the utility ratio (e.g., Luce & Raiffa, 1957)). Generally speaking, these methods use the desirability of outcomes to scale the consequences of each pair of decision outcome and true state. In the empirical example below, the correct and incorrect pass decisions (e.g., hiring and training costs) were perceived as respectively the most and the least preferred outcomes from the economic perspective of the company.

Qualified majority rule (QMR)

Quite often the collective decision rule g is given and not necessarily optimal. However, it might still be possible to improve the predictor-based selection process by controlling its optimal cutoff point R^*. Suppose now that a qualified majority rule (QMR) is employed, which is defined as follows:

$$g = \begin{cases} -1 & N(-1) \geq kn \\ 1 & \text{otherwise,} \end{cases}$$

where $N(-1)$ is the number of predictor tests failed by the applicant, n is the number of predictor tests, and k ($1/n \leq k \leq 1$ and kn is an integer) is the minimal proportion of predictor tests failed by the applicant necessary for the collective decision to be -1 (rejection of applicant). The parameter k represents the collective decision rule g, or the structure of the decision-making process. For instance, a simple majority rule $k = \frac{n+1}{2n}$ implies that an applicant is rejected if $N(-1) \geq \frac{n+1}{2}$ and accepted otherwise. It should be noticed that the assumption of a QMR is plausible because the optimal collective decision rule is always a qualified majority one, as shown in Ben-Yashar and Nitzan (1997). The problem we face is therefore:

$$\text{Max}_R \, \alpha U(1)\varphi(k/1) + (1-\alpha)U(-1)\varphi(k/-1).$$

Given the structure k of collective decision-making and the number n of predictor tests, the optimal cutoff point R^* for a multiple test is determined by the following necessary condition:

$$\frac{dp_1}{dR} = -Z\frac{dp_2}{dR}W, \tag{3}$$

where $Z = \frac{(1-\alpha)U(-1)}{\alpha U(1)}$ and $W = \frac{\frac{\partial \varphi(g/-1)}{\partial p_2}}{\frac{\partial \varphi(g/1)}{\partial p_1}} = \left(\frac{p_2}{1-p_1}\right)^{kn-1}\left(\frac{1-p_2}{p_1}\right)^{n-kn}$.

Following the line of reasoning given by Ben-Yashar and Nitzan (1998), the proof of the above assertion is given in Appendix A. If it is henceforth assumed that $\frac{d^2 p_1}{dR^2} < 0$, $\frac{d^2 p_2}{dR^2} < 0$, $\frac{\partial^2 \varphi(g/1)}{\partial p_1^2} < 0$ and $\frac{\partial^2 \varphi(g/-1)}{\partial p_2^2} < 0$, it is also shown in Appendix A that these conditions are sufficient to ensure that R^* is a maximum. In the empirical example below, these sufficient conditions will be examined for the situation that $f_1(x)$ and $f_2(x)$ are assumed to be normally distributed.

In a single test, it obviously holds that n, and thus k, is equal to 1 implying that $W = 1$. It follows then immediately from (3) that the optimal cutoff point R^* in this case is determined by the following necessary condition:

$$\frac{dp_1}{dR} = -Z\frac{dp_2}{dR}. \tag{4}$$

Note that (4) also follows immediately from (2) since $\varphi(g/1) = p_1$ and $\varphi(g/-1) = p_2$ for a single test. Also, it follows immediately from (4) that $\frac{d^2 p_1}{dR^2} < 0$, $\frac{d^2 p_2}{dR^2} < 0$ are sufficient conditions to ensure that R^* is a maximum.

The term Z which appears in equations (3) and (4) relates to the environmental characteristics of the decision-making process, viz., the prior that an applicant is suitable and the fixed utilities corresponding to the four possible outcomes for a predictor test. In fact, Z represents the quality of the selection environment. If state of nature 1 is superior to state of nature -1, the lower the Z, the higher the quality of this environment. $Z < 1$ represents an environment of relatively high quality. $Z > 1$ represents a relatively low-quality environment. $Z = 1$ represents a neutral environment. In other words, in this case there is no bias in favor of acceptance-rejection and pass-fail decisions of applicants in terms of respectively the priors of the two states of nature (i.e., $\alpha = (1-\alpha)$) and the net utilities corresponding to the two states of nature (i.e., $U(1) = U(-1)$).

The term W which appears in (3) and not in (4) is the ratio between the marginal contribution of a test's decisional skill to the collective probability of making a correct decision in states of natures –1 and 1. W depends on the three characteristics of the decision-making process: structure (collective decision rule), number of predictor tests composing a multiple test and a performance measure of its predictor tests which depends on their decisional skills. Note that when

BEST COPY AVAILABLE

$W = 1$ the effect of a marginal change in a test's decisional skill is identical under the two states of nature, that is, $\frac{\partial \varphi(g/1)}{\partial p_1} = \frac{\partial \varphi(g/-1)}{\partial p_2}$.

Relationship between Optimal Cutoff Points for Single and Multiple Tests

The optimal cutoff points for single and multiple tests in predictor-based selection are usually different. Whether or not the cutoff points for single tests are stricter than the cutoff points for multiple tests depend on the characteristics of the decision-making setting: the preferred decisional skills of the predictor tests, the number of predictor tests and the collective decision rule. Our main result specifies the condition that determines the relationship between the optimal cutoff points R^+ and R^* for single and multiple tests in predictor-based selection.

Theorem 1: $$R^* \gtrless R^+ \quad \Leftrightarrow \quad W \gtrless 1 \quad \Leftrightarrow \quad k \gtrless \lambda$$

where $$\lambda = \frac{1}{n} + \frac{n-1}{n} \cdot \frac{\ln \frac{p_1}{1-p_2}}{\beta_1 + \beta_2},$$

n is the fixed size of the number of predictor tests, $\beta_1 = \ln \frac{p_1}{1-p_1}$ and $\beta_2 = \ln \frac{p_2}{1-p_2}$.

The parameter λ can be interpreted as the bias/asymmetry of the tests' decisional skills.

Proof:

We first prove that $R^* \gtrless R^+ \quad \Leftrightarrow \quad W \gtrless 1$.

Let $h_1 = -\frac{dp_1}{dR} \cdot \frac{1}{Z}$, $h_2 = \frac{dp_2}{dR}$ and $h_3 = h_2 W$.

Notice that

$$\text{(i)} \quad h_1, h_2, h_3 > 0$$

$$\text{(ii)} \quad \frac{dh_1}{dR} > 0, \frac{dh_2}{dR} < 0$$

$$\text{(iii)} \quad \frac{\partial h_3}{\partial R} < 0.$$

By (3) and (4), R^* and R^+ are determined respectively by the equalities:

$$h_3(R^*) = h_1(R^*) \qquad \text{and} \qquad h_2(R^+) = h_1(R^+).$$

$$W \gtrless 1 \Leftrightarrow h_2(R^*) \lessgtr h_3(R^*) = h_1(R^*) \Leftrightarrow R^+ \lessgtr R^*.$$

We complete the proof by showing that $W \gtrless 1 \Leftrightarrow k \gtrless \lambda$.

Since $W = \left(\frac{p_2}{1-p_1}\right)^{kn-1} \left(\frac{1-p_2}{p_1}\right)^{n-kn}$,

$$W \gtrless 1 \Leftrightarrow (kn-1)\ln\frac{p_2}{1-p_1} + (n-kn)\ln\frac{1-p_2}{p_1} \gtrless 0$$

$$\Leftrightarrow kn\left(\ln\frac{p_2}{1-p_1}\cdot\frac{p_1}{1-p_2}\right) + n\ln\frac{1-p_2}{p_1} - \ln\frac{p_2}{1-p_1} \gtrless 0$$

$$\Leftrightarrow kn\left(\ln\frac{p_1}{1-p_1} + \ln\frac{p_2}{1-p_2}\right) \gtrless \ln\frac{p_2}{1-p_1} - n\ln\frac{1-p_2}{p_1}$$

$$\Leftrightarrow kn \gtrless \frac{\ln\frac{p_2}{1-p_1} + n\ln\frac{p_1}{1-p_2}}{\beta_1 + \beta_2} = \frac{(\beta_1 + \beta_2) + (n-1)\ln\frac{p_1}{1-p_2}}{\beta_1 + \beta_2}.$$

We have thus obtained that

$$W \gtrless 1 \Leftrightarrow kn \gtrless 1 + (n-1)\frac{\ln\frac{p_1}{1-p_2}}{\beta_1 + \beta_2} \quad \text{or} \quad W \gtrless 1 \Leftrightarrow k \gtrless \lambda.$$

Q.E.D.

The relationship between R^* and R^+ depends on the relationship between $\frac{\partial\varphi(g/-1)}{\partial p_2}$ and $\frac{\partial\varphi(g/1)}{\partial p_1}$ that depends on k. When $\frac{\partial\varphi(g/-1)}{\partial p_2} > \frac{\partial\varphi(g/1)}{\partial p_1}$, i.e., when $W > 1$, from the perspective of a multiple test there exists a relative advantage to an increase in tests' decisional skills in state of nature –1. This induces an increase in p_2 and a decrease in p_1 by setting the cutoff point R^* higher than R^+ (recall that $\frac{dp_1}{dR} < 0$ and $\frac{dp_2}{dR} > 0$). A similar argument can be used to rationalize the determination of R^* which is lower than R^+ when $W < 1$.

Alternatively, the relationship between R^* and R^+ depends on the relationship between k and λ. When $k > \lambda$, i.e., the structure of the multiple test is sufficiently lenient toward acceptance of applicants, the decision-making system reacts by setting a cutoff point higher than the one set on a single test, namely, by setting R^* which exceeds R^+. A similar argument can be used to rationalize the inequality $R^* < R^+$ when $k < \lambda$. Notice that the difference between R^* and R^+ is basically due to the interchangeability between R and k (Ben-Yashar & Nitzan, 1998).

To further clarify the intuition behind the theorem from a personnel selection perspective and, in particular, why $R^* \gtrless R^+ \Leftrightarrow k \gtrless \lambda$, let us first show that in a neutral environment λ approximates the optimal QMR for a multiple test consisting of a large number of predictor tests. In a neutral environment where $Z = 1$, the optimal QMR, k^*, is given by:

$$k^* = \frac{1}{2} + \frac{1}{2}\frac{\ln\frac{p_1(1-p_1)}{p_2(1-p_2)}}{\beta_1 + \beta_2},$$

which follows immediately from the optimal QMR in the general case that Z represents a bias/asymmetry in the environmental characteristics of the decision-making process (i.e., $Z \neq 1$):

$$k^* = \frac{1}{2}\left[1 + \frac{\ln\frac{1}{Z} + \frac{n}{2}\ln\frac{p_1(1-p_1)}{p_2(1-p_2)}}{n\frac{\beta_1+\beta_2}{2}}\right]. \qquad (5)$$

Appendix B provides a proof of the above assertion, which is a slightly modified version of the proof given in Ben-Yashar and Nitzan (1997).

Notice that

$$\lambda = \frac{1}{n} + \frac{n-1}{n}\cdot\frac{\ln\frac{p_1}{1-p_2}}{\beta_1+\beta_2}$$

$$= \frac{n+1}{2n} + \frac{n-1}{n}\left[\frac{\ln\frac{p_1}{1-p_2}}{\beta_1+\beta_2} - \frac{1}{2}\right]$$

$$= \frac{n+1}{2n} + \frac{n-1}{n}\cdot\frac{2\ln\frac{p_1}{1-p_2} - \ln\frac{p_1}{1-p_2} - \ln\frac{p_2}{1-p_1}}{2(\beta_1+\beta_2)}$$

$$= \frac{n+1}{2n} + \frac{n-1}{n}\cdot\frac{1}{2}\cdot\frac{\ln\frac{p_1(1-p_1)}{p_2(1-p_2)}}{\beta_1+\beta_2},$$

which converges to k^* for a sufficiently large n. When $p_1 = p_2$, $\lambda = 0.5$; that is, the optimal QMR is the simple majority rule.

When $p_1 < p_2$, $\lambda > 0.5$; that is, the optimal QMR favours the acceptance of applicants which is less likely to be the correct decision. If the given collective decision rule, k, implies a bias that optimally takes into account the difference between p_1 and p_2, i.e., $k = \lambda$, then from the perspective of a multiple test there is no incentive to set R and, in turn, p_1 and p_2 that differ from those set in a single test. In such a case $R^* = R^+$. If a collective decision rule k is faced that implies a bias in favour of selection of applicants which is stronger than the optimal bias corresponding to p_1 and p_2, i.e., $k > \lambda$, then an incentive exists to adjust p_1 and p_2 in order to eliminate the discrepancy between k and λ. The adjustment requires an increase of R which reduces p_1 and raises p_2, and therefore in such a case $R^* > R^+$. A similar argument can be used for the case $k < \lambda$, which completes our intuitive

explanation why $R^* \gtrless R^+ \Leftrightarrow k \gtrless \lambda$.

Disjunctive and Conjunctive Rules

A number of implications can be obtained from Theorem 1 in special cases of our model, that is, when specific assumptions are made regarding n, k, and the relationship between the endogenous decisional skills of predictor tests, p_1 and p_2.

As already noted, the structure of the decision-making system is represented by k, the minimal proportion of predictor tests in favour of alternative -1 (fail decision), necessary for the collective decision to be -1 (rejection of applicant). The following discussion, however, pertains to the structure of the decision-making system necessary for collectively reaching the decision 1 (acceptance of applicant). Hence $kn = n$ means that the collective rule is a *disjunctive* one. That is, if one predictor test decides in favour of alternative 1 (pass decision), then the collective decision is 1 (acceptance). $kn = 1$ means that the collective rule is a *conjunctive* one. That is, the collective decision is 1 (acceptance) only when an applicant is passed on all predictor tests. In fact, a conjunctive test can be interpreted as a case of multiple hurdles in personnel psychology.

By assumption, $p_1 > (1-p_2)$. Hence, for a conjunctive rule where $kn = 1$,

$$W = \left(\frac{p_2}{1-p_1}\right)^{kn-1}\left(\frac{1-p_2}{p_1}\right)^{n-kn} = \left(\frac{1-p_2}{p_1}\right)^{n-1} < 1.$$

By Theorem 1, in such a case $R^* < R^+$. In the extreme case of a disjunctive rule where $kn = n$, and using $p_2 > (1-p_1)$, $W = \left(\frac{p_2}{1-p_1}\right)^{n-1} > 1$ and, by Theorem 1, $R^* > R^+$.

The determination of optimal cutoff points for multiple tests takes into account the collective decision rule, k, and the interchangeability between k and the cutoff point R. No wonder then that for a disjunctive rule where the collective decision rule is most lenient toward acceptance of applicants, stricter cutoff points are set relative to the cutoff points set on single tests. In contrast, for a conjunctive rule, where the collective decision rule is least lenient toward acceptance of applicants, more tolerant cutoff points are set relative to the cutoff points set on single tests.

In the symmetric case where there is no bias in favor of acceptance-rejection and pass-fail decisions of applicants both in terms of the collective decision rule, $k = \frac{n+1}{2n}$, and in terms of the predictor cutoff which results in $p_1 = p_2$, the same cutoff points are set on single and multiple tests. Formally, since

$$W \gtrless 1 \Leftrightarrow \left(\frac{p_2}{1-p_1}\right)^{kn-1} \gtrless \left(\frac{p_1}{1-p_2}\right)^{n-kn} \Leftrightarrow \frac{p_2}{1-p_1} \gtrless \frac{p_1}{1-p_2} \Leftrightarrow p_1 \gtrless p_2 \qquad (6)$$

and $kn - 1 \gtrless n - kn \Leftrightarrow kn \gtrless \frac{n+1}{2}$, we obtain that if a simple majority rule is applied, that is, if $k = \frac{n+1}{2n}$ and the cutoff point yields identical decisional skills of the predictor tests, $p_1 = p_2$, as frequently assumed in the literature, then $W = 1$ and $\lambda = \frac{n+1}{2n} = k$. By Theorem 1, in such a case $R^{\bullet} = R^{+}$.

Finally, suppose that $[\alpha U(1) = (1\text{-}\alpha)U(\text{-}1)]$. In this symmetric situation (i.e., $Z = 1$), the optimal cutoff point R^{+} on the single test is set by maximizing its expected utility $[\alpha U(1)p_1 + (1\text{-}\alpha)U(\text{-}1)p_2]$ implying that its average decisional skill $\frac{p_1 + p_2}{2}$ is maximized. In general, R^{+} differs from $R^{\bullet}$. However, regardless of whether $R^{\bullet} > R^{+}$ or $R^{\bullet} < R^{+}$, in this symmetric situation there is always a tendency that the average decisional skills of the predictor tests of a multiple test are reduced relative to the average decisional skills of a single test.

Predictor-based Selection using the Assessment Center Method: An Illustration

To illustrate the comparison of the optimal cutoff points R^{+} and $R^{\bullet}$ set on single and multiple tests by using Theorem 1, the assessment center method is given as an empirical example. The term refers to a procedure for evaluating the performance of individuals for such purposes as selection or promotion of employees (e.g., Roos et al., 1997). In a typical assessment center the candidates applying for a job participate in a variety of exercises that enable them to demonstrate a particular skill, knowledge, ability, or competence, usually called job dimensions. These dimensions resemble the future professional practice as much as possible. The performance rating on each exercise is done by observers (called assessors) who are carefully trained in order for the method to be valid

and reliable. Comparing these ratings with a pre-established cutoff point, it is decided whether or not an applicant's performance on each specific exercise is satisfactorily enough to be passed. Then the assessors combine the pass-fail decisions on all the exercises and reach a collective decision for each applicant, that is, either accept or reject the applicant for the job.

In the current example data were available for a large company. The candidates applying for trainee positions in this company spent two days undergoing assessment of their managerial potential by the assesssment center method. The following 15 exercises were identified: Oral communication, planning and organization, written communication, analysis, reading skills, judgment, initiative, sensitivity, leadership, management identification, delegation, technical knowledge, reflection, trouble shooting, and presentation. The performance on each of the 15 exercises (i.e., the predictor tests i) was rated by a team of two carefully trained assessors on a 100-point scale running from 0 to 100. So, i was running from 1 to 15 and the predictor score x_i was running from 0 to 100.

Since the 15 teams of carefully trained assessors could be considered as approximately equally skilled in the area they had to rate the applicant's performance, the decisional skills of each of the 15 teams of assessors were assumed to be equal (i.e., homogeneous decisional skills), that is, $p_1^i = p_1$ and $p_2^i = p_2$ $(1 \le i \le 15)$. In other words, the conditional probabilities of correct pass-fail decisions for suitable and unsuitable applicants were assumed to be equal for each of the 15 teams of assessors. Furthermore, since the company did not have any prior information of the applicants, the a priori probabilities α and $(1-\alpha)$ of qualifying an applicant's true state (i.e., future job behavior) as respectively suitable ($s = 1$) or unsuitable ($s = 0$) were set equal. Hence, $\alpha = (1-\alpha) = 0.5$.

Using the lottery method described in Luce and Raiffa (1957), the positive net utility corresponding to a correct pass decision (i.e., $U(1)$) was perceived by the company from an economic perspective twice as large as the positive net utility corresponding to a correct fail decision (i.e., $U(-1)$). Hence, since the utility ratio $U(1)/U(-1) = 2$ and $\alpha = (1-\alpha) = 0.5$, it follows that $Z = 1/2$. Since $Z < 1$, we are thus dealing with environmental characteristics of the assessment center that can be characterized as being of relatively high quality.

In order to calculate the optimal cutoff point R^* by means of (3), given the collective decision rule k and number n of exercises, we finally still need to specify p_1 and p_2 as functions of R. It was assumed that the test score distributions $f_1(x_i)$ and $f_2(x_i)$ in the suitable and unsuitable group of applicants followed the normal distribution with mean μ_1 and μ_2 (with μ_2 lower than μ_1) and standard deviation σ_1 and σ_2, respectively. Based on a sample of 127 candidates (69 accepted

and 58 rejected) applying for trainee positions in the past, the parameters μ_1, μ_2, σ_1, and σ_2 were estimated in the following way.

First, for each candidate (both accepted and rejected ones) a composite predictor score x was calculated by taking his or her average over all 15 exercises. Henceforth, the composite predictor score will be denoted as X whereas future job performance will be denoted as the criterion variable Y. Next, for each selected applicant a criterion score y (i.e., applicant's supervisor rating of current job performance on a 100-point scale) was determined on the criterion variable Y. For the group of selected applicants the following statistics could now be computed for X and Y: the means μ_X and μ_Y, the standard deviations σ_X and σ_Y, and the correlation ρ_{XY} (i.e., validity coefficient) between X and Y. Using these statistics, we then computed for each rejected applicant the predicted criterion score $\hat{y}$ (i.e., future job behaviour if the applicant would have been selected) as a linear regression estimate on applicant's composite predictor score x:

$$\hat{y} = \mu_Y + \rho_{XY}(\sigma_X/\sigma_Y)(x-\mu_X). \tag{7}$$

Note that Brogden's utility model (1949) also assumed the above linear regression estimate for a single employee in which $\hat{y}$ then stands for the dollar value of an employee's performance.

Next, for each rejected applicant it was determined if he or she would have been classified as suitable or unsuitable by examining if applicant's predicted criterion score $\hat{y}$ was above or below a pre-established cutoff point $y_c = 55$ on the criterion variable Y. Also, each selected applicant was classified as suitable or unsuitable; that is, whether or not his or her criterion score y assigned by applicant's supervisor exceeded y_c. The mean and standard deviation of $f_1(x)$ and $f_2(x)$ could now be estimated straightforward yielding $\mu_1 = 72.83$, $\mu_2 = 47.51$, $\sigma_1 = 10.28$, and $\sigma_2 = 11.14$. The assumption of normality for $f_1(x)$ and $f_2(x)$ was tested using a Kolmogorov-Smirnov goodness-of-fit test. It turned out that the p-values were respectively 0.329 and 0.265, showing a satisfactory fit (significance level of 0.05) against the data.

Thus, using the customary notation $\Phi(\mu, \sigma)$ for the normal distribution with mean μ and standard deviation σ, the cumulative density is $\Phi(\mu_1, \sigma_1)$ for the suitable and $\Phi(\mu_2, \sigma_2)$ for the unsuitable applicants. It then follows that $p_1 = 1 - \Phi((R-\mu_1)/\sigma_1)$ (where $\Phi((R-\mu_1)/\sigma_1)$ now represents the lower tail probability of the standard normal distribution evaluated at the cutoff point R),

whereas $p_2 = \Phi((R-\mu_2)/\sigma_2)$.

Sufficient conditions for R^* and R^+ being a maximum

For given structure k, the two sufficient conditions $\frac{d^2 p_1}{dR^2} < 0$, $\frac{d^2 p_2}{dR^2} < 0$ ensuring that (2) has a maximum can now be examined. Let $\phi(z)$ refer to the standard normal density evaluated at z, it then follows that $\frac{dp_1}{dR} = -\phi((R-\mu_1)/\sigma_1)$ and that $\frac{dp_2}{dR} = \phi((R-\mu_2)/\sigma_2)$. Next, it follows that $\frac{d^2 p_1}{dR^2} = (1/\sigma_1)[(R-\mu_1)/\sigma_1]\phi((R-\mu_1)/\sigma_1)$ and $\frac{d^2 p_2}{dR^2} = -(1/\sigma_2)[(R-\mu_2)/\sigma_2]\phi((R-\mu_2)/\sigma_2)$. It can easily be verified that the two sufficient conditions for a maximum will be satisfied only in the case that $\mu_2 < R < \mu_1$. This constraint makes sense, since it implies that the mean μ_1 of the suitable applicants must be higher than R^*, whereas the mean μ_2 of the unsuitable applicants must be lower than R^*. In addition to the sufficient conditions $\frac{\partial^2 \varphi(g/1)}{\partial p_1^2} < 0$ and $\frac{\partial^2 \varphi(g/-1)}{\partial p_2^2} < 0$ for a maximum of (2), the constraint $\mu_2 < R < \mu_1$ will be examined below for concrete values of R^*.

Also, notice that the constraint $\mu_2 < R < \mu_1$ is perfectly consistent with the assumption that $(p_1+p_2)/2 > 0.5$, or, equivalently, that $[\Phi((R-\mu_2)/\sigma_2) - \Phi((R-\mu_1)/\sigma_1)] > 0$. This inequality holds because $\Phi((R-\mu_2)/\sigma_2) > 0.5$ and $\Phi((R-\mu_1)/\sigma_1) < 0.5$; that is, $[\Phi((R-\mu_2)/\sigma_2) - \Phi((R-\mu_1)/\sigma_1)] > 1$.

Relation between R^* and R^+ for given values of k and $n = 15$

R^+ was computed by inserting $\frac{dp_1}{dR} = -\phi((R-\mu_1)/\sigma_1)$, $\frac{dp_2}{dR} = \phi((R-\mu_2)/\sigma_2)$, and $Z = 0.5$ into (4) resulting in $R^+ = 57.22$. R^+ was computed numerically using a root finding procedure from the software package Mathematica (Wolfram, 1996).

In order to investigate the influence of more and less lenient collective rules on the optimal predictor cutoff, R^* was computed for $k = 3/15$, $k = 8/15$, and $k = 13/15$. Inserting first $k = 3/15$ and $n = 15$ into W and next W and $Z = 0.5$ into (3), and using again the root finding procedure from Mathematica (Wolfram, 1996), resulted in $R^* = 49.22$, $W = 0.157$, $\lambda = 0.226$, $p_1 = 0.989$, and $p_2 = 0.561$. So, verifying Theorem 1 for $k = 3/15 = 0.2$ results in:

$$R^{\bullet} = 49.22 < R^{+} = 57.22 \quad \Leftrightarrow \quad W = 0.157 < 1 \quad \Leftrightarrow \quad k = 0.2 < \lambda = 0.226.$$

As can be seen from the above result, $R^{\bullet} < R^{+}$ implying that a more tolerant cutoff point is set on the multiple test composed of 15 exercises relative to the cutoff point set on each single exercise. This result can be accounted for that the collective rule $k = 3/15$ is much less lenient toward selection of applicants than the simple majority rule since $kn = 3 < 8$ (i.e., (15+1)/2). This 'conjunctive like' character of the collective rule $k = 0.2$ also implies that p_1 is so large (and thus, p_2 so low) due to only selecting applicants from which we can be pretty sure that they will be qualified as suitable in their future job performance.

Next, for $k = 8/15 = 0.533$ (i.e., the simple majority rule), we obtained the following results: $R^{\bullet} = 60.31$, $W = 1.999$, $\lambda = 0.522$, $p_1 = 0.888$, and $p_2 = 0.875$. According to Theorem 1, a somewhat stricter cutoff point $R^{\bullet}$ is now set on the multiple test composed of 15 exercises relative to the cutoff point R^{+} set on each single exercise. This makes sense since the simple majority rule is more lenient toward selection of applicants than the collective rule $k = 3/15$. As a consequence of the more lenient character of the simple majority rule, p_1 and p_2 were respectively decreased and increased relative to the collective rule $k = 3/15$. It can easily be verified that the simple majority rule meets the requirement formulated in (6), since $W = 1.999 > 1 \Leftrightarrow p_1 = 0.888 > p_2 = 0.875$.

Finally, for $k = 13/15 = 0.867$, we obtained the following results: $R^{\bullet} = 70.99$, $W = 19.66$, $\lambda = 0.821$, $p_1 = 0.571$, and $p_2 = 0.982$. As can be verified from Theorem 1 (i.e., $W >> 1$), a much stricter cutoff point $R^{\bullet}$ is now set on the multiple test composed of 15 exercises relative to the cutoff point R^{+} set on each single exercise. This is because the collective rule $k = 13/15$ is much more lenient toward selection of applicants than the simple majority rule. This 'disjunctive like' character of the collective rule $k = 13/15$ also accounts for the finding that p_2 is so large (and thus, p_1 so low) since we only reject applicants from which we can be pretty sure that they would be qualified as unsuitable in their future job performance.

Relation between $R^{\bullet}$ and R^{+} for given value of $n = 15$

$R^{\bullet}$ and $k^{\bullet}$ will be determined dependently for $n = 15$ by maximizing simultaneously the multiple test's common expected utility and subsequently comparing $R^{\bullet}$ with R^{+} again. First k is written as function of R according to (5), then this function is inserted into (3) and solved for $R^{\bullet}$. In doing so, according to the definition of a QMR, kn must be rounded off to the next highest integer. Using again a root finding procedure from the software package Mathematica (Wolfram, 1996), yielded

the following results: $R^* = 58.40$, $k^* = 7/15 = 0.467$, $W = 1.306$, $\lambda = 0.462$, $p_1 = 0.920$, and $p_2 = 0.836$. As is clear from Theorem 1, a somewhat stricter cutoff point R^* is now set on the multiple test composed of 15 exercises relative to the cutoff point $R^+ = 57.22$ set on each single exercise. The optimal cutting points R^* and R^+, however, are nearly the same in case R and k are optimized simultaneously. Also, the optimal collective rule is only one exercise more lenient toward selection of applicants than the simple majority rule; that is, 9 out of 15 versus 8 out of 15 exercises must be passed at least for being accepted, respectively.

Observe that Type I error (i.e., $1\text{-}p_1 = 0.080$) is smaller than Type II error (i.e., $1\text{-}p_2 = 0.164$). This result is desirable from the economic perspective of the company since the probability of selecting applicants who turn out to be unsuitable in their future job performance should be lower than the probability of rejecting applicants who would have been suitable in their future job performance.

Examining the conditions sufficient for R^* and R^+ being a maximum

As is clear from the results, the sufficient condition $\mu_2 < R^* < \mu_1$ for (2) having a maximum was satisfied in all of the above examples. Also, it turned out that $\frac{\partial^2 \varphi(g/1)}{\partial p_1^2} < 0$ and $\frac{\partial^2 \varphi(g/\text{-}1)}{\partial p_2^2} < 0$ for all values of R^* in the above examples. Furthermore, a maximum was guaranteed for R^+ since $\mu_2 < R^+ < \mu_1$.

It should be mentioned that the sufficient condition of $\mu_2 < R^* < \mu_1$ did not hold for $n = 15$ and $k = 1/15$ (conjunctive rule), 2/15, 14/15, and 1 (disjunctive rule) since R^* turned namely out to be equal to respectively 40.10, 45.68, 74.18, and 78.83. The sufficient condition for (2) having a maximum, however, still appeared to hold due to the fact that the second derivative of (2) with respect to R was negative for these values of R^*. This finding can be accounted for that the second derivative of (2) with respect to R consists of the sum of four terms (see also Appendix A) of which either the second ($k = 1/15, 2/15$) or the fourth term ($k = 14/15, 1$) was now positive, whereas the other three terms still turned out to be negative for these values of R^*.

As an aside, it may be noted that the requirements of $W < 1$ and $W > 1$ for respectively a conjunctive and disjunctive rule were met since $W = 0.02$ for $k = 1/15$ (conjunctive rule) and $W = 95.23$ for $k = 1$ (disjunctive rule).

Concluding Remarks

This paper focuses on the comparison between the optimal cutoff points set on single and multiple tests in predictor-based selection. Since the characteristics of the two types of tests differ, these cutoff points that determine in our dichotomous choice model the decisional skills of the predictor tests are usually different. The relationship between them depends on the number of predictor tests composing a multiple test, on its collective decision rule, and on the tests' decisional skills. Our main result implies that the cutoff point for a multiple test is stricter than the cutoff point set on a single test, if the collective decision rule is sufficiently lenient toward acceptance of candidates applying for a job, as in the extreme case of a disjunctive rule. More generally, the structure of the decision-making process applies stricter cutoff points for selection of applicants if the marginal contribution of a test's decisional skill to the collective probability of rejecting unsuitable applicants is larger than its marginal contribution to the collective probability of accepting suitable applicants.

Our results are applied to compare the predictor cutoffs adopted in centralized selection systems and less informed decentralized selection systems. Clearly, decentralized predictor-based decision making in selection systems based on incomplete information can be improved. This is illustrated in the context of collective decision-making using the assessment center method by teams of assessors regarding the acceptance or rejection of candidates applying for trainee positions in a large company.

Two possible lines of future research arise from the present study. The first line is the extension to situations where it is not realistic to assume homogeneous decisional skills. Such is the case, for instance, if assessors are not equally skilled in the areas they have to rate applicants' performance. Loosening this assumption complicates matters since a vector of optimal predictor cutoff values must now be derived instead of one value for R^*.

A second possible line of research would be, following the classical Brogden-Cronbach-Gleser suggestion (Brogden, 1949; Cronbach & Gleser, 1965), to express the utility function rather as a function of the predictor cutoff R than as a threshold utility like in the present paper. For instance, analogous to Brogden's pioneering utility equation (1949), by expressing the utility for a single employee (i.e., the observed dollar value of an employee's job performance) as a linear regression on the composite predictor score (see also (7)). The choice of this utility function would be more in line with current models of personnel selection utility.

Acknowledgements

The authors are indebted to Saskia Klomp for providing the data for the empirical example and to Wim M.M. Tielen for his computational support.

References

Ben-Yashar, R., & Nitzan, S. (1997). The optimal decision rule for fixed size committees in dichotomous choice situations: The general result. *International Economic Review, 38*(1), 175-187.

Ben-Yashar, R., & Nitzan, S. (1998). Quality and structure of organizational decision making. *Journal of Economic Behavior and Organization, 36*, 521-534.

Ben-Yashar, R., & Nitzan, S. (2001). Investment criteria in single and multi-member economic organizations. *Public Choice, 109*, 1-13.

Boudreau, J.W. (1991). Utility analysis for decisions in human resource management. In M.D. Dunette and L.M. Hough (Eds.), *Handbook of industrial and organizational psychology* (Vol. 2, 2nd ed., pp. 621-745). Palo Alto, CA: Consulting Psychologists Press.

Brogden, H. (1949). When testing pays off. *Personnel Psychology, 2*, 171-183.

Chuang, D.T., Chen, J.J., & Novick, M.R. (1981). Theory and practice for the use of cut-scores for personnel decisions. *Journal of Educational Statistics, 6*, 129-152.

Cronbach, L.J., & Gleser, G.L. (1965). *Psychological tests and personnel decisions* (2nd ed.). Urbana, Ill.: University of Illinois Press.

Cronbach, L.J., & Snow, R.E. (1977). *Aptitudes and instructional methods: A handbook for research on interactions*. New York: Irvington Publishers, Inc.

De Corte, W. (1994). Utility analysis for the one-cohort selection decision with a probationary period. *Journal of Applied Psychology, 79*, 402-411.

De Corte, W. (1998). Estimating and maximizing the utility of sequential selection decisions with a probationary period. *British Journal of Mathematical and Statistical Psychology, 51*, 101-121.

DeGroot, M.H. (1970). *Optimal statistical decisions*. New York: McGraw-Hill.

Hambleton, R.K., Swaminathan, H., & Rogers, H.J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage.

Huynh, H. (1977). Two simple classes of mastery scores based on the beta-binomial model. *Psychometrika, 42*, 601-608.

Lehmann, E.L. (1959). *Testing statistical hypotheses* (3rd ed.). New York: Macmillan.

Lewis, C., & Sheehan, K. (1990). Using Bayesian decision theory to design a computerized mastery test. *Applied Psychological Measurement, 14*, 367-386.

Luce, R.D., & Raiffa, H. (1957). *Games and decisions*. New York: John Wiley and Sons.

Milkovich, G.T., & Boudreau, J.W. (1997). *Human resource management* (8th ed.). Chicago, IL: Richard D. Irwin.

Petersen, N.S. (1976). An expected utility model for 'optimal' selection. *Journal of Educational Statistics, 4*, 333-358.

Raju, N.S., Steinhaus, S.D., Edwards, J.E., & Lehessio, J. (1991). A logistic regression model for personnel selection. *Applied Psychological Measurement, 15*, 139-152.

Roos, J., Roos, G., Dragonetti, N.C., & Edvinsson, L. (1997). *Intellectual capital: Navigating the new business landscape*. London: Macmillan Press Ltd.

Sackett, P.R. & Roth, L. (1996). Multi-stage selection strategies: A Monte Carlo investigation of effects on performance and minority hiring. *Personnel Psychology, 49*, 549-572.

Taylor, H.C., & Russell, J.T. (1939). The relationship of validity coefficients to the practical effectiveness of tests in selection: Discussion and tables. *Journal of Applied Psychology, 23*, 565-578.

van der Linden, W.J. (1981). Using aptitude measurements for the optimal assignment of subjects to treatments with and without mastery scores. *Psychometrika, 46*, 257-274.

van der Linden, W.J. (1987). The use of test scores for classification decisions with threshold utility. *Journal of Educational Statistics, 12*, 62-75.

van der Linden, W.J., & Vos, H.J. (1996). A compensatory approach to optimal selection with mastery scores. *Psychometrika, 61*, 155-172.

Vos, H.J. (1997a). Simultaneous optimization of quota-restricted selection decisions with mastery scores. *British Journal of Mathematical and Statistical Psychology, 50*, 105-125.

Vos, H.J. (1997b). A simultaneous approach to optimizing assignments with mastery scores. *Multivariate Behavioral Research, 32*, 403-433.

Vos, H.J. (2001). A minimax procedure in the context of sequential testing problems in psychodiagnostics. *British Journal of Mathematical and Statistical Psychology, 54*, 139-159.

Wolfram, S. (1996). *The Mathematica book* (3rd ed.). Cambridge: Cambridge University Press.

Appendix A

Theorem: Given the structure k of decision making and the number n of predictor tests, the optimal cutoff point R^* for a multiple test is determined by the following necessary (first-order) condition:

$$\frac{dp_1}{dR} = -Z\frac{dp_2}{dR}W,$$

whereas the following (second-order) conditions are sufficient for R^* being a maximum:

$$\frac{d^2p_1}{dR^2} < 0, \quad \frac{d^2p_2}{dR^2} < 0, \quad \frac{\partial^2\varphi(g/1)}{\partial p_1^2} < 0, \quad \frac{\partial^2\varphi(g/-1)}{\partial p_2^2} < 0.$$

Proof: The solution R^* of the optimization problem in (2) satisfies the following necessary (first-order) condition:

$$\alpha U(1)\frac{\partial\varphi(g/1)}{\partial p_1}\frac{dp_1}{dR} + (1-\alpha)U(-1)\frac{\partial\varphi(g/-1)}{\partial p_2}\frac{dp_2}{dR} = 0.$$

When the collective decision rule g is a QMR and kn is an integer, and using $N(-1) + N(1) = n$, it holds:

$$g = \begin{cases} -1 & N(-1) \geq kn \Leftrightarrow n - N(1) \geq kn \Leftrightarrow N(1) \leq n - kn \\ 1 & \text{otherwise} \Leftrightarrow N(1) > n - kn \Leftrightarrow N(1) \geq n + 1 - kn, \end{cases}$$

and therefore

$$\varphi(g/1) = \sum_{i=n+1-kn}^{n}\binom{n}{i}(p_1)^i(1-p_1)^{n-i},$$

$$\varphi(g/-1) = \sum_{i=kn}^{n}\binom{n}{i}(p_2)^i(1-p_2)^{n-i}.$$

It follows:

$$\frac{\partial \varphi(g/1)}{\partial p_1} = \frac{\partial \sum_{i=n+1-kn}^{n} \binom{n}{i} (p_1)^i (1-p_1)^{n-i}}{\partial p_1}$$

$$= \sum_{i=n+1-kn}^{n} \binom{n}{i} i(p_1)^{i-1}(1-p_1)^{n-i} - \sum_{i=n+1-kn}^{n} \binom{n}{i} (p_1)^i (n-i)(1-p_1)^{n-i-1}.$$

Substituting $j = i - 1$ in the right-hand side expression, and using the obvious identities

$$\binom{n}{i} i = n\binom{n-1}{i-1} \quad \text{and} \quad (n-1)\binom{n}{i} = n\binom{n-1}{i}, \text{ we get}$$

$$\frac{\partial \varphi(g/1)}{\partial p_1} = \sum_{j=n-kn}^{n-1} n\binom{n-1}{j} (p_1)^j (1-p_1)^{(n-1)-j} - \sum_{i=n-1-kn}^{n} n\binom{n-1}{i} (p_1)^i (1-p_1)^{(n-1)-i}$$

$$= n\binom{n-1}{n-kn} (p_1)^{n-kn} (1-p_1)^{kn-1}.$$

In a similar way we find that

$$\frac{\partial \varphi(g/\text{-}1)}{\partial p_2} = n\binom{n-1}{kn-1} (p_2)^{kn-1} (1-p_2)^{n-kn}.$$

By substituting $\frac{\partial \varphi(g/1)}{\partial p_1}$ and $\frac{\partial \varphi(g/\text{-}1)}{\partial p_2}$ in the first-order condition, we obtain

$$\alpha U(1) \frac{dp_1}{dR} n\binom{n-1}{n-kn} (p_1)^{n-kn} (1-p_1)^{kn-1}$$

$$+ (1-\alpha) U(-1) \frac{dp_2}{dR} n\binom{n-1}{kn-1} (p_2)^{kn-1} (1-p_2)^{n-kn} = 0.$$

Hence,

$$\frac{dp_1}{dR} = -\frac{(1-\alpha)U(-1)}{\alpha U(1)}\frac{dp_2}{dR}\left(\frac{p_2}{1-p_1}\right)^{kn-1}\left(\frac{1-p_2}{p_1}\right)^{n-kn} = -Z\frac{dp_2}{dR}W.$$

The necessary condition for (2) having a maximum is also sufficient if the second derivative of (2) with respect to R is negative, implying that the following inequality must hold:

$$\alpha U(1)\frac{\partial^2\varphi(g/1)}{\partial p_1^2}\left(\frac{dp_1}{dR}\right)^2 + \alpha U(1)\frac{\partial\varphi(g/1)}{\partial p_1}\frac{d^2p_1}{dR^2}$$

$$+ (1-\alpha)U(-1)\frac{\partial^2\varphi(g/-1)}{\partial p_2^2}\left(\frac{dp_2}{dR}\right)^2 + (1-\alpha)U(-1)\frac{\partial\varphi(g/-1)}{\partial_2}\frac{d^2p_2}{dR^2} < 0.$$

The assumed inequalities

$$\frac{d^2p_1}{dR^2} < 0, \quad \frac{d^2p_2}{dR^2} < 0, \quad \frac{\partial^2\varphi(g/1)}{\partial p_1^2} < 0, \quad \frac{\partial^2\varphi(g/-1)}{\partial p_2^2} < 0$$

ensure that the above (second-order) condition is satisfied. This completes the proof.

Appendix B

Theorem: For given n and R (and thus, p_1 and p_2), the optimal qualified majority rule (QMR), k^*, is given by:

$$k^* = \frac{1}{2}\left[1 + \frac{\ln\frac{1}{Z} + \frac{n}{2}\ln\frac{p_1(1-p_1)}{p_2(1-p_2)}}{n\frac{\beta_1+\beta_2}{2}}\right].$$

Proof: For any decision profile $a = (a_1,\ldots, a_n)$ in Ω, consider the partition of the predictor tests into $B(a)$ and $D(a)$ such that $i \in B(a)$ if $a_i = 1$ and $i \in D(a)$ if $a_i = -1$. Note that $|D(a)| = N(-1)$ and $|B(a)| = N(1)$, implying that $|D(a)| + |B(a)| = n$. Furthermore, let $t(a/1)$ and $t(a/-1)$ denote the conditional probabilities to obtain the decision profile $a = (a_1,\ldots, a_n)$, given that the applicant is suitable (should be accepted) or unsuitable (should be rejected). In other words,

$$t(a/1) = \prod_{i\in B(a)} p_1\left(\frac{a_i+1}{2}\right) \prod_{i\in D(a)} (1-p_1)\left(\frac{1-a_i}{2}\right) \text{ and}$$

$$t(a/-1) = \prod_{i\in D(a)} p_2\left(\frac{1-a_i}{2}\right) \prod_{i\in B(a)} (1-p_2)\left(\frac{a_i+1}{2}\right).$$

For a given collective decision rule g, it holds:

$$\varphi(g/1) = \sum_{a\in A(g/1)} t(a/1) \qquad \text{and} \qquad \varphi(g/-1) = \sum_{a\in A(g/-1)} t(a/-1).$$

It follows that the solution of the maximization problem in (2) on which we focus is also the solution of the problem:

$$\text{Max}_g \ \alpha U(1) \sum_{a\in A(g/1)} t(a/1) + (1-\alpha)U(-1) \sum_{a\in A(g/-1)} t(a/-1).$$

The partition of Ω to $A(g/1)$ and $A(g/-1)$ is determined by the control variable of the above maximization problem, namely by the collective decision rule g. Consider the following condition: for any profile of decisions a, $a \in \Omega$,

$$g(a) = \begin{cases} 1 & \text{if} \quad \alpha U(1)t(a/1) > (1-\alpha)U(-1)t(a/-1) \\ -1 & \text{if} \quad \alpha U(1)t(a/1) \leq (1-\alpha)U(-1)t(a/-1). \end{cases}$$

This condition ensures that g maximizes the multiple test's common expected utility in the problem above and therefore in (2) over the set of all collective rules. In other words, a sufficient condition for the optimality of g^* is that it partitions Ω, such that

$$g^*(a) = -1 \Leftrightarrow \alpha U(1)t(a/1) \leq (1-\alpha)U(-1)t(a/-1)$$

$$\Leftrightarrow \frac{\alpha}{1-\alpha}\frac{U(1)}{U(-1)} \prod_{i \in B(a)} p_1\left(\frac{a_i+1}{2}\right) \prod_{i \in D(a)} (1-p_1)\left(\frac{1-a_i}{2}\right) \leq$$

$$\prod_{i \in D(a)} p_2\left(\frac{1-a_i}{2}\right) \prod_{i \in B(a)} (1-p_2)\left(\frac{a_i+1}{2}\right)$$

$$\Leftrightarrow \ln\frac{1}{Z} + \sum_{i \in B(a)} \left(\ln\frac{p_1}{1-p_2}\right)\frac{a_i+1}{2} \leq \sum_{i \in D(a)} \left(\ln\frac{p_2}{1-p_1}\right)\frac{1-a_i}{2}$$

$$\Leftrightarrow \ln\frac{1}{Z} + \sum_{i=1}^{n} \left(\ln\frac{p_1}{1-p_2}\right)\frac{a_i+1}{2} - \sum_{i=1}^{n} \left(\ln\frac{p_2}{1-p_1}\right)\frac{1-a_i}{2} \leq 0$$

$$\Leftrightarrow \ln\frac{1}{Z} + \sum_{i=1}^{n} \frac{1}{2}\left(\ln\frac{p_1}{1-p_1} + \ln\frac{p_2}{1-p_2}\right) a_i + \sum_{i=1}^{n} \frac{1}{2} \ln\frac{p_1(1-p_1)}{p_2(1-p_2)} \leq 0.$$

Rearranging terms, and substituting $\beta_1 = \ln\frac{p_1}{1-p_1}$ and $\beta_2 = \ln\frac{p_2}{1-p_2}$, it follows:

$$g^*(a) = -1 \quad \Leftrightarrow \quad -\sum_{i=1}^{n} a_i \geq \frac{\ln\frac{1}{Z} + \frac{n}{2}\ln\frac{p_1(1-p_1)}{p_2(1-p_2)}}{\frac{\beta_1+\beta_2}{2}}$$

$$\Leftrightarrow \quad |D(a)| - |B(a)| \geq \frac{\ln\frac{1}{Z} + \frac{n}{2}\ln\frac{p_1(1-p_1)}{p_2(1-p_2)}}{\frac{\beta_1+\beta_2}{2}}.$$

Using $|D(a)| + |B(a)| = n$, it follows:

$$g^*(a) = -1 \quad \Leftrightarrow \quad |D(a)| \geq \frac{n}{2}\left[1 + \frac{\ln\frac{1}{Z} + \frac{n}{2}\ln\frac{p_1(1-p_1)}{p_2(1-p_2)}}{n\frac{\beta_1+\beta_2}{2}}\right].$$

Hence, and using $N(-1) = |D(a)|$, it finally follows by the definition of a QMR:

$$k^* = \frac{1}{2}\left[1 + \frac{\ln\frac{1}{Z} + \frac{n}{2}\ln\frac{p_1(1-p_1)}{p_2(1-p_2)}}{n\frac{\beta_1+\beta_2}{2}}\right].$$

Q.E.D.

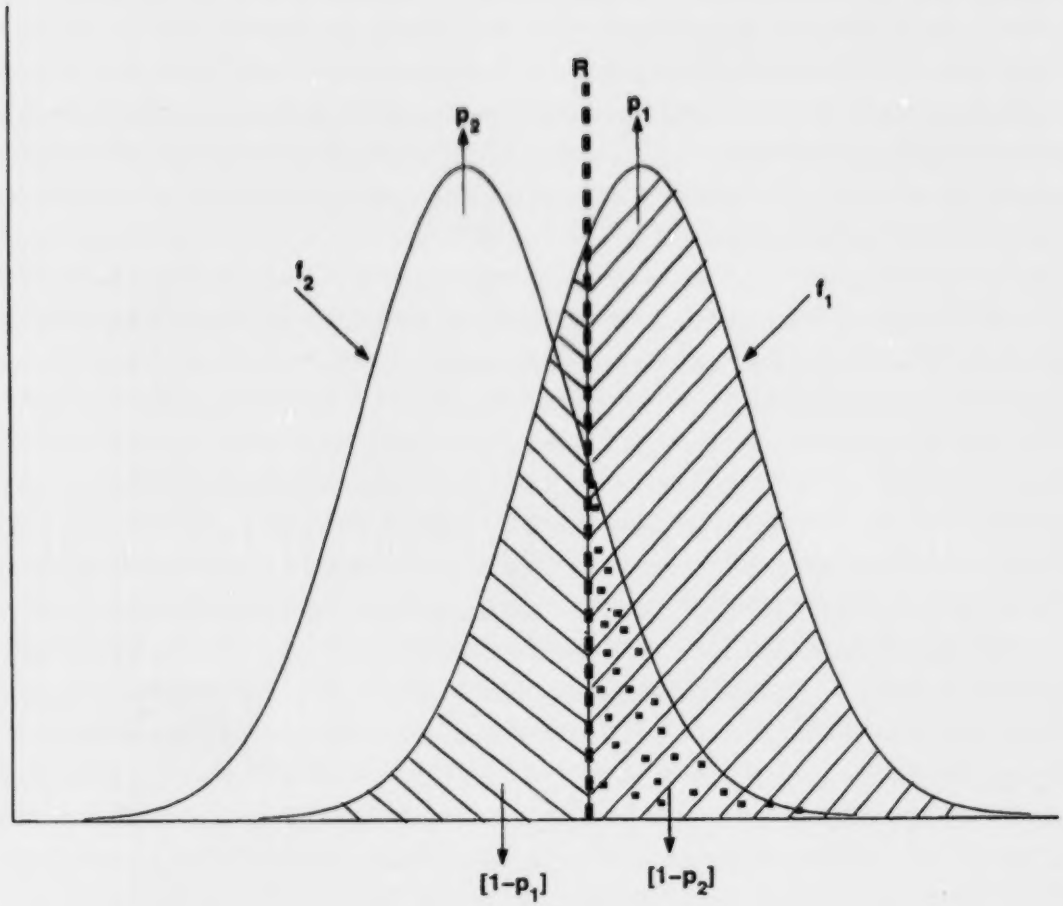

Figure 1

Titles of Recent Research Reports from the Department of Educational Measurement and Data Analysis. University of Twente, Enschede, The Netherlands.

RR-01-04 R. Ben-Yashar, S. Nitzan & H.J. Vos, *Optimal Cutoff Points in Single and Multiple Tests for Psychological and Educational Decision Making*

RR-01-03 R.R. Meijer, *Outlier Detection in High-Stakes Certification Testing*

RR-01-02 R.R. Meijer, *Diagnosing Item Score Patterns using IRT Based Person-Fit Statistics*

RR-01-01 W.J. van der Linden & H. Chang, *Implementing Content Constraints in Alpha-Stratified Adaptive testing Using a Shadow test Approach*

RR-00-11 B.P. Veldkamp & W.J. van der Linden, *Multidimensional Adaptive Testing with Constraints on Test Content*

RR-00-10 W.J. van der Linden, *A Test-Theoretic Approach to Observed-Score equating*

RR-00-09 W.J. van der Linden & E.M.L.A. van Krimpen-Stoop, *Using Response Times to Detect Aberrant Responses in Computerized Adaptive Testing*

RR-00-08 L. Chang & W.J. van der Linden & H.J. Vos, *A New Test-Centered Standard-Setting Method Based on Interdependent Evaluation of Item Alternatives*

RR-00-07 W.J. van der linden, *Optimal Stratification of Item Pools in a-Stratified Computerized Adaptive Testing*

RR-00-06 C.A.W. Glas & H.J. Vos, *Adaptive Mastery Testing Using a Multidimensional IRT Model and Bayesian Sequential Decision Theory*

RR-00-05 B.P. Veldkamp, *Modifications of the Branch-and-Bound Algorithm for Application in Constrained Adaptive Testing*

RR-00-04 B.P. Veldkamp, *Constrained Multidimensional Test Assembly*

RR-00-03 J.P. Fox & C.A.W. Glas, *Bayesian Modeling of Measurement Error in Predictor Variables using Item Response Theory*

RR-00-02 J.P. Fox, *Stochastic EM for Estimating the Parameters of a Multilevel IRT Model*

RR-00-01 E.M.L.A. van Krimpen-Stoop & R.R. Meijer, *Detection of Person Misfit in Computerized Adaptive Tests with Polytomous Items*

RR-99-08 W.J. van der Linden & J.E. Carlson, *Calculating Balanced Incomplete Block Designs for Educational Assessments*

RR-99-07 N.D. Verhelst & F. Kaftandjieva, *A Rational Method to Determine Cutoff Scores*

RR-99-06 G. van Engelenburg, *Statistical Analysis for the Solomon Four-Group Design*

RR-99-05 E.M.L.A. van Krimpen-Stoop & R.R. Meijer, *CUSUM-Based Person-Fit Statistics for Adaptive Testing*

RR-99-04 H.J. Vos, *A Minimax Procedure in the Context of Sequential Mastery Testing*

RR-99-03 B.P. Veldkamp & W.J. van der Linden, *Designing Item Pools for Computerized Adaptive Testing*

RR-99-02 W.J. van der Linden, *Adaptive Testing with Equated Number-Correct Scoring*

RR-99-01 R.R. Meijer & K. Sijtsma, *A Review of Methods for Evaluating the Fit of Item Score Patterns on a Test*

RR-98-16 J.P. Fox & C.A.W. Glas, *Multi-level IRT with Measurement Error in the Predictor Variables*

RR-98-15 C.A.W. Glas & H.J. Vos, *Adaptive Mastery Testing Using the Rasch Model and Bayesian Sequential Decision Theory*

RR-98-14 A.A. Béguin & C.A.W. Glas, *MCMC Estimation of Multidimensional IRT Models*

RR-98-13 E.M.L.A. van Krimpen-Stoop & R.R. Meijer, *Person Fit based on Statistical Process Control in an AdaptiveTesting Environment*

RR-98-12 W.J. van der Linden, *Optimal Assembly of Tests with Item Sets*

RR-98-11 W.J. van der Linden, B.P. Veldkamp & L.M. Reese, *An Integer Programming Approach to Item Pool Design*

RR-98-10 W.J. van der Linden, *A Discussion of Some Methodological Issues in International Assessments*

...

Research Reports can be obtained at costs, Faculty of Educational Science and Technology, University of Twente, TO/OMD, P.O. Box 217, 7500 AE Enschede, The Netherlands.

faculty of

EDUCATIONAL SCIENCE AND TECHNOLOGY

A publication by
The Faculty of Educational Science and Technology of the University of Twente
P.O. Box 217
7500 AE Enschede
The Netherlands

BEST COPY AVAILABLE

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

NOTICE

Reproduction Basis

X	This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.
	This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").